UNITED STATES CIVIL SERVICE COMMISSION
Office of the Executive Director
Interagency Advisory Group
1900 E Street, N.W.
Washington, D.C. 20415

Minutes of the IAG Committee on Personnel Research
May 22, 1978

The meeting was chaired by Dr. Kenneth R. Brown, Chief, Applied Psychology Section, Personnel Research and Development Center, Bureau of Policies and Standards. The purpose of the meeting was to present reports and provide opportunity for discussion about topics of interest to the Federal personnel research community. The agenda included a discussion of draft Uniform Guidelines requirements for the calculation of adverse impact, an update on Computerized Tailored Testing, and a presentation on the job relatedness and validity of the PACE written test.

Mr. John Curnow, Chief, Workforce Analysis and Statistics Division, BPMIS, and Mr. John Kraft, Chief, Test Services Section, PRDC, discussed different aspects of race, sex and ethnic group data collection. Mr. Curnow informed the group of activities relating to the resurveying of the on-board Federal workforce for the purpose of promulgating the OMB A-46 standard. A pilot test is being proposed to test the alternative means of collection (self vs. visual identification), and alternative questions (combined or separate race and ethnic designators). It is hoped that the pilot test will be carried out in the Fall of 1978 with implementation by the Fall of 1979. It was pointed out that this work does not directly involve implementation of the draft guidelines, and that interim guidance for agencies relating to guidelines data collection hopefully will be ready within three months of guidelines issuance.

Mr. John Kraft discussed work relating to race, sex and ethnic group data collection on applicants for Federal employment. Pilot studies on PACE applicants carried out in Washington and regions worked out well, with only a minor discrepancy between visual and self identification. (Visual identification was carried out only on black-white-other and male-female categories, not on ethnicity). Based on these results it has been concluded that collecting data in the examination room will work. Data collection efforts on exams other than PACE are being carried out, including mail-in responses with the Mid-level examination in the Boston region. Adverse reaction to data collection has come mainly from groups such as the American Jewish Committee, American Jewish Congress and Anti-Defamation League, which are of the opinion that data may be misused and that the Federal government should not track people in this manner. Other groups feel that data should be collected on other groups, such as the white ethnics of Europe. Mr. Kraft also informed the group of the activities involving comparing the Federal workforce with labor market data, stating that a match is often difficult and that the problems with such efforts are many. He also stated that a training package

relating to agency data collection should be completed by mid-summer and that the main concern regarding such data collection is protection of data - insuring that it is used for program evaluation purposes only and not selection of individuals.

Dr. Vern Urry and Mr. Paul Croll, Research Section, PRDC presented an update on Computerized Tailored Testing (CTT). Dr. Urry described the basic process of CTT in which the computer draws on a bank of items, flashes the selected item on a screen, and receives an answer from the examinee. After each response the computer makes an ability estimate and standard error estimate. Questions continue until the standard error reaches a predetermined level. For example, if the tailored tests for all examinees were terminated at a standard error of .21, we know that the reliability of the ability estimates would be .96 and that the correlation between true and estimated ability would be .98. Dr. Urry pointed out that the primary obstacles to CTT have been overcome. For example, it can be shown that with CTT a test need contain only 20% as many items as a standard paper and pencil test in order to attain the same reliability. It can also be shown that CTT can actually result in a major cost saving over time.

Mr. Croll presented a functional comparison of current and CTT systems. It was shown that CTT subsumes many of the activities of current testing systems including validation. Capabilities for additional functions are being developed. These entail immediate feedback, career appraisal and counseling, and estimates of the dollar value of current and alternative selection strategies. In addition, it was pointed out that CTT will provide the capability to agencies of conducting in-house testing either with CSC items or their own items, as well as transportability between agencies of validation procedures and question banks. To accomplish these activities will require the acquisition of a number of self contained computer terminals (currently about $2,000 each).

Mr. Richard McKillip, Chief, and Dr. Marvin Trattner, Research Section, PRDC presented an update on Test 500 criterion related research studies and provided information on future research. Mr. McKillip briefly described the four criterion related validity studies which have been undertaken to provide additional empirical support for the construct validity of the test. Studies were carried out using claims examiners in two bureaus of the Social Security Administration, Internal Revenue Officers, and Customs Inspectors. CODAP was used to help develop criteria and select the samples (it was desired to select a group of people in each occupation who performed similar duties). Criteria were specially developed for each study including job information tests, work samples, supervisory ratings, supervisory rankings, and training success. Reliabilities of criteria, where available, ranged from .57 to .86. Ten of the twelve validity coefficients were statistically significant. In the case of those that were not, the criteria were not really appropriate because of low reliability and because supervisors ordinarily do not observe the work of subordinates. The conclusion (especially keeping in mind that the obtained validity coefficients are conservative estimates of validity) was that those scoring higher on Test 500 do perform better on the job, providing empirical support to the construct validity of PACE.

Dr. Trattner discussed the proposed use of a synthetic validity approach for demonstrating the validity of PACE. Synthetic validity involves the calculation of a validity coefficient for an occupation without obtaining predictor and criterion scores for the occupation. The procedure would involve the delineation of the major duties of PACE jobs (it is anticipated that 50-100 duties should cover most PACE occupations adequately) and determining the validity of the test for measuring success in the duties. The synthetic validity coefficient for an occupation is then calculated by having the validity coefficients weighted for importance by subject matter experts (SME's). Dr. Trattner pointed out that agencies will probably not be as closely involved in synthetic validity research as they were in the criterion related validity studies. However, many SME's will be needed to generate duties and rate importance. It is anticipated that the research will deal with the most populous PACE occupations and those agencies with the greatest numbers in those occupations.